ON A FORMULA FOR THE PRODUCT-MOMENT COEFFICIENT OF ANY ORDER OF A NORMAL FREQUENCY DISTRIBUTION IN ANY NUMBER OF VARIABLES.

By L. ISSERLIS, D.Sc.

1. In *Biometrika*, Vol. XI, Part III, I have shown that for a normal frequency distribution in four variables, if

$$p_{xyzt} = \underset{x\ y\ z\ t}{SSSS} \{n_{xyzt}\, xyzt\}/N$$

denotes the product-moment coefficient of the distribution about the means of the four variables and q_{xyzt} is the *reduced* moment, i.e.

$$q_{xyzt} = p_{xyzt}/\sigma_x\sigma_y\sigma_z\sigma_t,$$

then

$$q_{xyzt} = r_{xy}r_{zt} + r_{yz}r_{xt} + r_{zx}r_{yt} \quad \text{.........................(1).}$$

In this result any two or more variables may be made identical leading to a variety of results for moment coefficients of distributions containing fewer than four variables but of total order four, for example identifying t with x we obtain

$$q_{x^2yz} = r_{yz} + 2r_{xy}r_{xz} \quad \text{.................................(2),}$$

and putting $y = z = t = x$ we find $q_{x^4} = 3$; of course $q_{xy} = r_{xy}$ and q_{x^4} is merely β_2.

I suggested that (1) was probably capable of generalisation, and I now propose to prove a general theorem which gives immediately the value of the mixed moment coefficient of any order in each variable for a normal frequency distribution in any number of variables.

2. Consider a normal distribution, total population N. Let $N_{12\ldots n}$ denote the frequency of the group in which the characters differ by $x_1, x_2, \ldots x_n$ from the mean values for the whole population and let

$$p_{1^{l_1}2^{l_2}\ldots n^{l_n}} = S\,(N_{12\ldots n}x_1^{l_1}x_2^{l_2}\ldots x_n^{l_n})/N \quad \text{.....................(3),}$$

denote the moment coefficient of the most general kind about the mean values of the characters. The corresponding reduced moment will be

$$q_{1^{l_1}2^{l_2}\ldots n^{l_n}} = p_{1^{l_1}2^{l_2}\ldots n^{l_n}}/\sigma_1^{l_1}\sigma_2^{l_2}\ldots \sigma_n^{l_n} \quad \text{........................(4).}$$

Then for *normal distributions*,

$$\text{if } n \text{ be } odd, \quad q_{12\ldots n} = 0 \quad \text{...................................(5),}$$

$$\text{and if } n \text{ be } even, \quad q_{12\ldots n} = S\,(r_{ab}r_{cd}\ldots r_{hk}) \quad \text{.................(6),}$$

where the summation on the right-hand side extends to every possible selection of $n/2$ pairs $ab, cd, \ldots hk$, that can be formed out of the n suffixes $1, 2, 3, \ldots n$; equation (1) is thus a particular case of (6).

Equation (6) is the theorem it is proposed to prove. The value of $q_{1^{l_1}2^{l_2}\ldots n^{l_n}}$ is at once found for given numerical values of the indices $l_1, l_2, \ldots l_n$ by writing down (5) for $l_1 + l_2 + \ldots + l_n$ variables and identifying the values of l_1 of them with that of the first and so on.

For example if we require the value of $q_{1^2 2^2 3^2}$ we commence with,

$$\begin{aligned} q_{123456} &= S\,(r_{ab} r_{cd} r_{ef}) \\ &= r_{12}\,(r_{34} r_{56} + r_{35} r_{46} + r_{36} r_{45}) + r_{13}\,(r_{24} r_{56} + r_{25} r_{46} + r_{26} r_{45}) \\ &\quad + r_{14}\,(r_{23} r_{56} + r_{25} r_{36} + r_{26} r_{35}) + r_{15}\,(r_{23} r_{46} + r_{24} r_{36} + r_{26} r_{34}) \\ &\quad + r_{16}\,(r_{23} r_{45} + r_{24} r_{35} + r_{25} r_{34}) \quad \text{......(7).} \end{aligned}$$

Identifying 4 with 1, 5 with 2 and 6 with 3 we find at once

$$q_{1^2 2^2 3^2} = 1 + 2r_{12}^2 + 2r_{23}^2 + 2r_{31}^2 + 8r_{12} r_{23} r_{31} \quad \text{......(8).}$$

3. We note first that q_{1^n} which in the more usual notation for distributions in one variable is $\mu_n/\mu_2^{n/2}$ is known to have the value $1.3.5 \ldots (n-1)$ when n is even. As regards $S\,(r_{ab} r_{cd} \ldots r_{hk})$, if all the n variables are made identical, each term becomes unity and the number of terms is the same as the number of ways of breaking up an even number (n) of objects into ($n/2$) pairs. This last number is clearly

$$\frac{n!}{2!\,\overline{n-2}!}\,\frac{\overline{n-2}!}{2!\,\overline{n-4}!} \cdots \frac{4!}{2!\,2!} \Big/ (n/2)!$$

which also reduces to $1.3.5 \ldots (n-1)$; thus equation (6) is correct for this particular case.

Secondly let us consider the value of $q_{1^{n-1}2}$. The mean value of x_2 for a given value of x_1 is $r_{12}\sigma_2 x_1/\sigma_1$, let

$$x_2 = r_{12}\frac{\sigma_2}{\sigma_1} x_1 + X_2.$$

Then the distribution of X_2 for a given value of x_1 is itself normal and its kth moment is zero for an odd k and

$$1.3.5 \ldots (k-1)\,({}_1\sigma_2)^{k/2}$$

for an even k where ${}_1\sigma_2$ is the standard deviation of 2 within the x_1 array so that ${}_1\sigma_2^2 = (1 - r_{12}^2)\,\sigma_2^2$.

$$\begin{aligned} q_{1^{n-1}2} &= \frac{1}{\sigma_1^{n-1}\sigma_2} \text{ Mean value } (x_1^{n-1} x_2) \\ &= \frac{1}{\sigma_1^{n-1}\sigma_2} \text{ Mean } \left\{ x_1^{n-1} \text{ Mean } \left(r_{12}\frac{\sigma_2}{\sigma_1} x_1 + X_2 \right) \right\} \\ &= r_{12} q_{1^n} = 1.3.5 \ldots (n-1)\, r_{12} \quad \text{......(9).} \end{aligned}$$

The method employed in the original proof of equation (1) is not convenient for generalisation and we will now prove the equation

$$q_{1234} = r_{12} r_{34} + r_{13} r_{24} + r_{14} r_{23}$$

by the method that leads to the general case.

Putting as above

$$x_2 = r_{12}\frac{\sigma_2}{\sigma_1} x_1 + X_2,$$

$$x_3 = r_{13}\frac{\sigma_3}{\sigma_1} x_1 + X_3,$$

$$x_4 = r_{14}\frac{\sigma_4}{\sigma_1} x_1 + X_4,$$

we have

$$p_{1234} = \text{Mean of } (x_1 x_2 x_3 x_4)$$
$$= \text{Mean of } \{x_1 \text{ (Mean of } x_2 x_3 x_4 \text{ for a given value of } x_1)\}$$
$$= \text{Mean of } \left[x_1 \left\{\text{Mean of } \left(r_{12}\frac{\sigma_2}{\sigma_1}x_1 + X_2\right)\left(r_{13}\frac{\sigma_3}{\sigma_1}x_1 + X_3\right)\left(r_{14}\frac{\sigma_4}{\sigma_1}x_1 + X_4\right)\right\}\right].$$

Now for normal distributions (and if the original distribution is normal, so is that within the x_1 array), Mean $X_2 = 0$, Mean $X_2X_3X_4 = 0$, while

$$\text{Mean } X_2X_3 = ({}_1\sigma_2)({}_1\sigma_3)\,{}_1r_{23}$$
$$= \sqrt{1 - r_{12}^2}\,\sigma_2\sqrt{1 - r_{13}^2}\,\frac{\sigma_3 (r_{23} - r_{12}r_{13})}{\sqrt{1 - r_{12}^2}\sqrt{1 - r_{13}^2}}$$
$$= (r_{23} - r_{12}r_{13})\,\sigma_2\sigma_3 \quad \text{..............(10).}$$

Hence

$$p_{1234} = \text{Mean of } \left[x_1 \left\{r_{12}r_{13}r_{14}\,\sigma_2\sigma_3\sigma_4\frac{x_1^3}{\sigma_1^3} + r_{12}\sigma_2\frac{x_1}{\sigma_1}(r_{34} - r_{13}r_{14})\,\sigma_3\sigma_4 + r_{13}\sigma_3\frac{x_1}{\sigma_1}(r_{24} - r_{12}r_{14})\,\sigma_2\sigma_4 + r_{14}\sigma_4\frac{x_1}{\sigma_1}(r_{23} - r_{12}r_{13})\,\sigma_2\sigma_3\right\}\right],$$

or dividing by $\sigma_1\sigma_2\sigma_3\sigma_4$,

$$q_{1234} = r_{12}r_{13}r_{14}q_{1^4} + q_{1^2}\{r_{12}(r_{34} - r_{13}r_{14}) + r_{13}(r_{24} - r_{12}r_{14})\} + r_{14}(r_{23} - r_{12}r_{13})$$
$$= r_{12}r_{34} + r_{23}r_{14} + r_{14}r_{23},$$

since $q_{1^2} = 1$ and $q_{1^4} = 3$. Thus our formula is established for the case of four variables.

4. We will establish the case for n variables by induction, and it will be convenient to denote by ${}_1q_{234\ldots n}$ the value of the reduced product-moment coefficient for the variables 2, 3, 4, ... n *within the x_1 array* so that

$${}_1q_{234\ldots n} = \frac{\text{Mean value of } (X_2X_3\ldots X_n)}{({}_1\sigma_2)({}_1\sigma_3)\ldots({}_1\sigma_n)},$$

where $X_2, X_3, \ldots X_n$ denote as before the deviations of the variables from their means within the x_1 array. Of course when n is even,

${}_1q_{234\ldots n}$ is zero since $n - 1$ is now odd.

Let n be even and assume that our formula has been proved true for all even values of n up to $n - 2$ inclusive, then

$$p_{123\ldots n} = \text{Mean } (x_1x_2x_3\ldots x_n)$$
$$= \text{Mean}\left\{x_1\left(r_{12}\sigma_2\frac{x_1}{\sigma_1} + X_2\right)\left(r_{13}\sigma_3\frac{x_1}{\sigma_1} + X_3\right)\ldots\left(r_{1n}\sigma_n\frac{x_1}{\sigma_1} + X_n\right)\right\}$$
$$= r_{12}r_{13}\ldots r_{1n}\sigma_2\sigma_3\ldots\sigma_n \text{ Mean } (x_1^n)/\sigma_1^{n-1}$$
$$+ S\{(r_{1a}r_{1b}r_{1c}\ldots)(\sigma_a\sigma_b\sigma_c\ldots)\text{ Mean } (X_\alpha X_\beta)\}\text{ Mean } (x_1^{n-2})/\sigma_1^{n-3}$$
$$+ S\{(r_{1a}r_{1b}r_{1c}\ldots)(\sigma_a\sigma_b\sigma_c\ldots)\text{ Mean } (X_\alpha X_\beta X_\gamma X_\delta)\}\text{ Mean } (x_1^{n-4})/\sigma_1^{n-5}$$
$$+ \ldots$$
$$+ S\{r_{1a}\sigma_a \text{ Mean } (X_\alpha X_\beta \ldots X_\gamma)\}\text{ Mean } (x_1^2)/\sigma_1 \quad \text{..............(11),}$$

the summations in each line extending to all possible permutations of the suffixes 2, 3, 4, ... n. The last line for example being

$$\frac{\text{Mean}\,(x_1^2)}{\sigma_1}\{r_{12}\sigma_2\,\text{Mean}\,(X_3X_4\ldots X_n)+r_{13}\sigma_3\,\text{Mean}\,(X_2X_4X_5\ldots X_n)+\ldots$$
$$+\,r_{1n}\sigma_n\,\text{Mean}\,(X_2X_3\ldots X_{n-1})\}.$$

Now we have seen that Mean $(X_2X_3)=(r_{23}-r_{12}r_{13})\,\sigma_2\sigma_3$. Similarly,

$$\begin{aligned}\text{Mean}\,(X_2X_3X_4X_5)&=({}_1\sigma_2)\,({}_1\sigma_3)\,({}_1\sigma_4)\,({}_1\sigma_5)\,({}_1q_{2345})\\&=({}_1\sigma_2)\,({}_1\sigma_3)\,({}_1\sigma_4)\,({}_1\sigma_5)\,[({}_1r_{23})\,({}_1r_{45})+({}_1r_{35})\,({}_1r_{24})+({}_1r_{25})\,({}_1r_{34})]\\&=(r_{23}-r_{12}r_{13})\,(r_{45}-r_{14}r_{15})+(r_{35}-r_{13}r_{15})\,(r_{24}-r_{12}r_{14})\\&\qquad+(r_{25}-r_{12}r_{15})\,(r_{34}-r_{13}r_{14}),\end{aligned}$$

and our assumption of the truth of equation (6) up to $(n-2)$ variables will enable us to write down the mean value of every product of X's occurring in (11).

Dividing by $\sigma_1\sigma_2\ldots\sigma_n$ we have, remembering that Mean x_1^n/σ_1^n is $1.3.5\ldots(n-1)$

$$\begin{aligned}q_{123\ldots n}&=(r_{12}r_{13}\ldots r_{1n})\,1.3.5\ldots(n-1)\\&\quad+S\,\{r_{1a}r_{1b}r_{1c}\ldots(r_{\alpha\beta}-r_{1\alpha}r_{1\beta})\}\,1.3.5\ldots(n-3)\\&\quad+S\,\{r_{1a}r_{1b}r_{1c}\ldots S'\,[(r_{\alpha\beta}-r_{1\alpha}r_{1\beta})\,(r_{\gamma\delta}-r_{1\gamma}r_{1\delta})]\}\,1.3.5\ldots(n-5)\\&\quad+\ldots\\&\quad+S\,\{r_{1a}S'\,[(r_{\alpha\beta}-r_{1\alpha}r_{1\beta})\,(r_{\gamma\delta}-r_{1\gamma}r_{1\delta})\,(r_{\epsilon\rho}-r_{1\epsilon}r_{1\rho})\ldots]\}\,.1\end{aligned}\qquad\text{.......(12)},$$

where S' refers to permutations of $\alpha\beta\gamma\ldots$ only, and S to permutations of all the suffixes $a, b, c, \ldots \alpha, \beta, \gamma \ldots$, i.e. all the suffixes 2, 3, 4, ... n.

It is clear that when the right-hand member of (12) is completely expanded no terms can survive which contain as a factor more than one correlation coefficient with suffix unity. This is easily verified in simple cases, and if in the general case a term $r_{1a}^{\lambda}r_{bc}r_{dc}\ldots$ survived, this term would reduce to r_{1a}^{λ} when we identified the characters a, 2, 3, ... n, which contradicts the value $1.3.5\ldots(n-1)\,r_{11}$ we have already found for it (equation (9)).

The value of the right-hand member is therefore easily found by neglecting all terms containing more than one such factor.

Hence on the assumption that (5) is true for all values of n up to $(n-2)$ we find

$$q_{123\ldots n}=S\,\{r_{1a}S'\,(r_{\alpha\beta}r_{\gamma\delta}r_{\epsilon\rho}\ldots)\},$$

but this is exactly the formula we wished to establish for it is obvious that $S\,(r_{ab}r_{cd}\ldots r_{hk})$ where $abc\ldots k$ is a permutation of $12\ldots n$ is equivalent to

$$S\,\{r_{1a}S'\,(r_{\alpha\beta}r_{\gamma\delta}\ldots)\}$$

where $a, \alpha, \beta, \gamma\ldots$ is a permutation of 2, 3, 4, ... n. Thus our formula which has been proved true for 4 variables is seen by induction to be true in general.

5. Formula (6) can be exhibited as a multiple definite integral: Let Δ denote the determinant whose kth row consists of the elements

$$(r_{1k},\,r_{2k},\,\ldots\,r_{k-1,k},\,\ldots\,r_{k+1,k},\,\ldots\,r_{nk})$$

and let Δ_{hk} denote the cofactor of the element in the hth row and kth column.

Let $$\chi^2 = \Sigma\left(\frac{\Delta_{hh}}{\Delta}\frac{x_h^2}{\sigma_h^2} + 2\Sigma\,\Delta_{hk}\frac{x_h x_k}{\sigma_h \sigma_k}\right),$$

and $$z = \frac{1}{(2\pi)^{\frac{n}{2}}\sigma_1\sigma_2\ldots\sigma_n\sqrt{\Delta}}\,e^{-\frac{1}{2}\chi^2},$$

then $$\int_{-\infty}^{+\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} x_1 x_2 \ldots x_n z\,dx_1 dx_2\ldots dx_n = S\,(r_{ab} r_{cd}\ldots r_{uv})\quad\ldots\ldots(13),$$

where $a, b, c, d, \ldots u, v$ are the suffixes $1, 2, 3, \ldots n$ in any possible order.

It is clear that (13) will enable us to write down the value of the multiple integral $\int_n Pe^{-Q}dx_1 \ldots dx_n$ where P is any polynomial in $x_1, x_2, \ldots x_n$ on Q a positive quadratic form.

In fact, let $\Sigma a_{pp}x_p^2 + 2\Sigma a_{pq}x_p x_q$, $(a_{pq} = a_{qp})$ be a positive. definite, quadratic form, then

$$W = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} x_1^{a_1} x_2^{a_2}\ldots x_n^{a_n} \exp - \tfrac{1}{2}\left(\Sigma a_{pp}x_p^2 + 2\Sigma a_{pq}x_p x_q\right) dx_1 dx_2 \ldots dx_n$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}}\sigma_1\sigma_2\ldots\sigma_n\sqrt{\Delta}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{x_1^{a_1}x_2^{a_2}\ldots x_n^{a_n}}{\sigma_1^{a_1}\sigma_2^{a_2}\ldots\sigma_n^{a_n}}$$

$$\exp - \frac{1}{2\Delta}\left(\Sigma\Delta_{pp}\frac{x_p^2}{\sigma_p^2} + 2\Sigma\Delta_{pq}\frac{x_p x_q}{\sigma_p\sigma_q}\right) dx_1 dx_2 \ldots dx_n$$

$= \Sigma\,[r_{ab} r_{cd} \ldots r_{hk}]$ where $abc \ldots hk$ is any permutation of the $a_1 + a_2 + \ldots + a_n$ suffixes of which a_1 are equal to 1, a_2 are equal to 2 and so on.

Let D denote the determinant of the quadratic form and D_{pq} the cofactor of a_{pq} the two multiple integrals will be identical if

$$1 = \sigma_1^2\sigma_2^2\ldots\sigma_{p-1}^2\sigma_{p+1}^2\ldots\sigma_n^2\Delta D_{pp},$$

$$r_{pq} = \sigma_1^2\sigma_2^2\ldots\sigma_p\sigma_q\ldots\sigma_n^2\Delta D_{pq}.$$

Hence $r_{pq}^2 = [D_{pq}]^2/D_{pp}D_{qq}$ and $\sigma_p^2 = D_{pp}/D$ while $\Delta = D^{n-1}/D_{11}D_{22}\ldots D_{nn}$,

so that $$W = \frac{(2\pi)^{\frac{n}{2}}}{D^{\frac{m}{2}+\frac{1}{2}}}\,\Sigma D_{ab}D_{cd}\ldots D_{hk}\quad\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(13'),$$

where $a, b, \ldots h, k$ is a permutation as above, and $m = a_1 + a_2 + \ldots + a_n$ is even. $W = 0$ when m is odd.

As an illustration of this result:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(Mx^2y^2z^2 + Nx^2yz)$$

$$\exp - \tfrac{1}{2}\left(ax^2 + by^2 + cz^2 + 2fyz + 2gzx + 2hxy\right) dx\,dy\,dz$$

$$= \frac{(2\pi)^{3/2}}{\Delta^{7/2}}M\,(8FGH + 2AF^2 + 2BG^2 + 2CH^2) + \frac{(2\pi)^{3/2}}{\Delta^{5/2}}N\,(2GH + AF),$$

where A, B, C, F, G, H are the cofactors of a, b, c, f, g, h in

$$\Delta = \begin{vmatrix} a, & h, & g \\ h, & b, & f \\ g, & f, & c \end{vmatrix}.$$

A cognate result is discussed by Mr Arthur Black in the *Transactions of the Cambridge Philosophical Society**. Black's integral is $\int_n Ve^{-U}dx_1 \dots dx_n$ where V and U are any *quadratic* functions, the only restriction on U being that it should be essentially positive. Other particular cases have been dealt with in the paper previously quoted, and for the case of two variables several results are given by Mr H. E. Soper†.

For reference we add a table of values of the reduced product-moment coefficients that occur frequently in formulae for probable errors and similar work.

$$q_{1^4} = 3.$$
$$q_{1^3 2} = 3r_{12}.$$
$$q_{1^2 2^2} = 1 + 2r_{12}^2.$$
$$q_{1^2 23} = r_{23} + 2r_{12}r_{13}.$$
$$q_{1^6} = 15.$$
$$q_{1^5 2} = 15r_{12}.$$
$$q_{1^4 2^2} = 3 + 12r_{12}^2.$$
$$q_{1^3 2^3} = 9r_{12} + 6r_{12}^3.$$
$$q_{1^3 2^2 3} = 3\,(r_{13} + 2r_{23}r_{12} + 2r_{13}r_{12}^2).$$
$$q_{1^4 23} = 3\,(r_{23} + 12r_{12}r_{13}).$$
$$q_{1^2 2^2 3^2} = 1 + 2r_{23}^2 + 2r_{31}^2 + 2r_{12}^2 + 8r_{12}r_{23}r_{31}.$$
$$q_{1^8} = 105, \quad q_{1^7 2} = 105r_{12}, \quad q_{1^6 2^2} = 15\,(6r_{12}^2 + 1).$$
$$q_{1^5 2^3} = 15\,(4r_{12}^3 + 3r_{12}).$$
$$q_{1^4 2^4} = 3\,(8r_{12}^4 + 24r_{12}^2 + 3).$$
$$q_{1^\lambda 23} = 1.3 \dots \lambda - 1\,(r_{23} + \lambda r_{12}r_{13}). \qquad \lambda \text{ even}.$$
$$q_{1^\lambda 2^2 3} = 1.3.5 \dots \lambda\,[(\lambda - 1)\,r_{12}^2 r_{13} + r_{13} + 2r_{12}r_{23}]. \qquad \lambda \text{ odd}.$$

For the case of two variables we add the following formula which is easily proved by the methods employed in this paper.

$$q_{1^u 2^v} = \psi\,(u + v)\,r^v + \binom{v}{2}\psi\,(2)\,\psi\,(u + v - 2)\,r^{v-2}\,(1 - r^2) + \binom{v}{4}\psi\,(4)\,\psi\,(u + v - 4)\,r^{v-4}\,(1 - r^2)^2 + \dots‡$$

the series terminating. Here

$$\psi\,(2m) = 1.3.5 \dots (2m - 1)$$

and

$$\binom{v}{m} = \frac{v\,(v - 1) \dots (v - m + 1)}{m!}.$$

* Vol. XVI, 1898, pp. 219—227.

† *Biometrika*, vol. IX, p. 101.

‡ This is virtually the formula (xxxii) employed by H. E. Soper, *l.c.a.* corrected for some misprints.